Optimal ε-biased sets with just a little randomness

Cristopher Moore* Alexander Russell†

April 19, 2013

Abstract

Subsets of $\mathbb{F}_2^n$ that are ε-*biased*, meaning that the parity of any set of bits is even or odd with probability ε close to $1/2$, are powerful tools for derandomization. A simple randomized construction shows that such sets exist of size $O(n/\varepsilon^2)$, and known deterministic constructions achieve sets of size $O(n/\varepsilon^3)$, $O(n^2/\varepsilon^2)$, and $O((n/\varepsilon^2)^{5/4})$. Rather than derandomizing these sets completely in exchange for making them larger, we attempt a partial derandomization while keeping them small, constructing sets of size $O(n/\varepsilon^2)$ with as few random bits as possible. The naive randomized construction requires $O(n^2/\varepsilon^2)$ random bits. We give two constructions. The first uses Nisan's space-bounded pseudorandom generator to partly derandomize a folklore probabilistic construction of an error-correcting code, and requires $O(n \log(1/\varepsilon))$ bits. Our second construction requires $O(n \log(n/\varepsilon))$ bits, but is more elementary; it adds randomness to a Legendre symbol construction on Alon, Goldreich, Håstad, and Peralta, and uses Weil sums to bound high moments of the bias.

1 Introduction

Derandomization is the art of replacing random choices with deterministic ones. In many cases, we can accomplish this by finding explicit constructions of combinatorial objects that "look random" in some sense. In particular, say a set $S \subseteq \mathbb{F}_2^n$ fools a function f if

$$\left| \mathop{\mathbb{E}}_{x \in S} f(x) - \mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n} f(x) \right| \leq \varepsilon \,,$$

for some small ε. If there are families of sets of polynomial size that we can construct in polynomial time, such that for each constant c we can fool every

*moore@santafe.edu, Santa Fe Institute and Department of Computer Science, University of New Mexico

†acr@cse.uconn.edu, Department of Computer Science and Engineering, University of Connecticut

$f \in \mathsf{TIME}(n^c)$ with ε sufficiently small, then every polynomial-time randomized algorithm can be derandomized and $\mathsf{P} = \mathsf{BPP}$.

In a number of applications, even sets that fool linear functions are useful [1]. In $\mathbb{F}_2^n$, any such function is the parity of some subset of x's coordinates. Let $x \in \mathbb{F}_2^n$ and let $T \subseteq [n]$. The parity of the bits of x indexed by T is

$$f_T(x) = \sum_{i \in T} x_i \,.$$

We say that S is *ε-biased* if, for all $T \neq \varnothing$,

$$\left| \Pr_{x \in S} [f_T(x) = 0] - \Pr_{x \in S} [f_T(x) = 1] \right| \leq \varepsilon \,.$$

Equivalently, if we identify $\mathbb{F}_2^n$ with $\{\pm 1\}^n$ in the natural way, then

$$\left| \mathop{\mathbb{E}}_{x \in S} \phi_T(x) \right| \leq \varepsilon \quad \text{where} \quad \phi_T(x) = \prod_{i \in T} x_i \,.$$

This is the same as saying that χ_S, the characteristic function of S, has a nearly flat Fourier spectrum: if we normalize it to $(1/|S|)\chi_S$, it has no coefficients greater than ε in absolute value. As a consequence, sampling a function on S gives a good approximation of its expectation if its Fourier spectrum has bounded ℓ_1 norm. In addition, ε-biased sets are important building blocks in other pseudorandom constructions; for instance, if $\varepsilon = n^{-c}$ for $c > 0$, then an ε-biased set is also approximately $O(\log n)$-wise independent.

There is a nice duality between ε-based sets and linear error-correcting codes [3]. Given an ε-biased set S, the truth table of each parity function $f_T(x)$ is a string in $\mathbb{F}_2^{|S|}$. Each such string is nearly balanced, with Hamming weight between $(1-\varepsilon)|S|/2$ and $(1+\varepsilon)|S|/2$. The set of parity functions has rank n, so an ε-biased set $S \in \mathbb{F}_2^n$ yields a $(|S|, n, d)$ code, i.e., a code of length $|S|$, rank n, and distance $d = (1-\varepsilon)|S|/2$. As a consequence, we can lower bound the size of an ε-biased set using sphere-packing arguments. As long as ε is not too small, and in particular if $\varepsilon = 1/\mathrm{poly}(n)$, this gives $|S| = \Omega(n/(\varepsilon^2 \log \varepsilon^{-1}))$ [2].

This lower bound is essentially tight, since we can construct an ε-biased set by choosing $O(n/\varepsilon^2)$ elements of $\mathbb{F}_2^n$ uniformly and independently. Equivalently, a random error-correcting code meets the Gilbert-Varshamov bound with high probability. Of course, this requires $n|S| = O(n^2/\varepsilon^2)$ random bits. Under the reasonable assumption that $\mathsf{TIME}(2^{O(n)}) \not\subseteq \mathsf{SPACE}(2^{O(n)})$, this construction can be generically derandomized [4]. But, as always, we are interested in derandomized constructions that work even in the absence of complexity assumptions.

Starting with [1], several deterministic constructions have been discovered, yielding ε-biased sets of size polynomial in n and $1/\varepsilon$. Depending on how ε scales with n, the best known constructions [2, 5] yield sets of size $O(n/\varepsilon^3)$, $O(n^2/\varepsilon^2)$, and $O((n/\varepsilon^2)^{5/4})$. The construction of [5] is especially notable; it applies Bezout's theorem from algebraic geometry, and achieves a set whose size is the 5/4 power of the optimum.

Tradeoffs between randomness and the quality of a combinatorial object is a classic topic in theoretical computer science. Here, we explore a different part of the randomness-size plane. Rather than reducing the amount of randomness to zero at the cost of making the set larger, we ask how much randomness we need to construct a set of optimal size, or equivalently what spaces we can succinctly describe that are guaranteed to contain at least one ε-biased set of optimal size.

Specifically, we give two randomness-efficient constructions of ε-biased sets of size $O(n/\varepsilon^2)$. While neither construction offers a witness that the resulting set is indeed ϵ-biased, both succeed with probability arbitrarily close to 1. The first uses $O(n \log(1/\varepsilon))$ random bits, using Nisan's space-bounded generator to partly derandomize a construction of random error-correcting codes. The second uses $O(n \log(n/\varepsilon))$ bits but is more elementary, and has a pleasant algebraic flavor: it works by "re-randomizing" a construction in [2] involving the Legendre symbol, and we use Weil sums to bound high moments of the bias. Note that if $\varepsilon = n^{-c}$ for $c > 0$, then in both constructions the number of random bits we need is much smaller than the set itself.

2 A random error-correcting code and Nisan's generator

First we review a simple folklore construction of a random linear error-correcting code whose distance is very close to half its length; this corresponds to an ε-biased set using the duality mentioned above. This construction already does noticeably better than the naive one, using $O(|S|) = O(n/\varepsilon^2)$ random bits. We then derandomize it further using a standard space-bounded pseudorandom generator, reducing the number of random bits to $O(n \log(1/\varepsilon))$.

Let $m \geq n$ and consider the finite field $\mathbb{F}_{2^m}$. We can identify each $x \in \mathbb{F}_2^n$ with an element of $\mathbb{F}_{2^m}$ in a way that preserves the additive structure of $\mathbb{F}_2^n$ by setting all but the last n bits to zero. For any fixed $\alpha \in \mathbb{F}_m$, we can then define a set of codewords in $\mathbb{F}_2^n \times \mathbb{F}_{2^m}$,

$$C_\alpha = \left\{ w_x = \big(x, \alpha x\big) \mid x \in \mathbb{F}_2^n \right\} .$$

Since multiplication by α is a linear function, C_α is closed under addition, making it a linear code. It has rank n and length $n + m$. We will show that, if $\alpha \in \mathbb{F}_{2^m}$ is uniformly random and m/n is sufficiently large, then C_α has distance $(1-\varepsilon)(n+m)/2$ with high probability, in which case it corresponds to an ε-biased set of size $n + m$. Equivalently, the Hamming weight $|w_x| = |x| + |\alpha x|$ of every nonzero codeword is at least $(1 - \varepsilon)(n + m)/2$.

For each nonzero $x \in \mathbb{F}_2^n$, αx is uniformly random in $\mathbb{F}_{2^m}$ since α is. Let $\delta \leq 1/2$. By the union bound, the probability that there is an $x \neq 0$ such that $|w_x| \leq \delta(n + m)$ is at most

$$P = 2^{-m} \sum_{k,j : k+j \leq \delta(n+m)} \binom{n}{k} \binom{m}{j} ,$$

where we sum over $k = |x|$ and $j = |\alpha x|$. This sum has at most n^2 terms, and the summand is maximized when $k = \delta n$ and $j = \delta m$, so

$$P \leq n^2\, 2^{-m} \binom{n}{\delta n} \binom{m}{\delta m} \leq n^2\, 2^{-m}\, \mathrm{e}^{h(\delta)(n+m)}\,, \tag{1}$$

where $h(\delta) = -\delta \ln \delta - (1-\delta)\ln(1-\delta)$ denotes the entropy function.

Now if $\delta = (1-\varepsilon)/2$, the Taylor series gives

$$h(\delta) \leq \ln 2 - \frac{\varepsilon^2}{2}\,,$$

and (1) becomes

$$P \leq n^2\, \mathrm{e}^{n \ln 2 - (\varepsilon^2/2)(n+m)}\,.$$

If we set

$$m = A\frac{n}{\varepsilon^2}$$

for some constant $A > 2\ln 2$, then $P = 2^{-\Omega(n)}$. Thus C_α has distance $(1-\varepsilon)(n+m)/2$ with high probability, giving an ε-biased set of size $n + m = O(n/\varepsilon^2)$.

To choose α uniformly would take $m = O(n/\varepsilon^2)$ random bits. However, we can do better by applying a pseudorandom generator for space-bounded computation. First, let us modify the construction somewhat, using $t = m/n = O(1/\varepsilon^2)$ blocks of n bits each. Rather than choosing α from $\mathbb{F}_{2^m}$, we write $\alpha = (\alpha_1, \ldots, \alpha_t)$ where $\alpha_i \in \mathbb{F}_{2^n}$ for each i. We then define each codeword as a concatenation of $t+1$ blocks,

$$C_\alpha = \left\{ w_x = \big(x, \alpha_1 x, \alpha_2 x, \ldots, \alpha_t x\big) \mid x \in \mathbb{F}_{2^n} \right\}.$$

If the α_i are uniformly random in $\mathbb{F}_{2^n}$ then so is $\alpha_i x$, and the probability that any w_x has Hamming weight less than $(1-\varepsilon)(n+m)/2$ is $2^{-\Omega(n)}$ just as before.

Now note that, for each $x \in \mathbb{F}_{2^n}$, there is a branching program B_x with states $\{0, \ldots, n+m\}$ that takes $\alpha_1, \ldots, \alpha_t$ as input and computes the total Hamming weight of w_x. Its initial state is $|x|$, on the ith step it reads α_i and increments its state by the weight of $\alpha_i x$, and it accepts if $|w_x| \geq (1-\varepsilon)(n+m)/2$. Our goal is to fool B_x with a pseudorandom sequence of tn bits, in such a way that the probability distribution of its final state has a total variation distance $o(2^{-n})$ from the distribution induced by uniformly random α_i. Taking a union bound over all x, the probability that any B_x rejects, i.e., that any w_x has Hamming weight less than $(1-\varepsilon)(n+m)/2$, is then $o(1)$ just as if the α_i were uniform. In that case, C_α is again an error-correcting code of distance $(1-\varepsilon)(n+m)/2$ with high probability.

We do this with Nisan's pseudorandom generator for space-bounded computation. Say that $f : \{0,1\}^\ell \to \{0,1\}^{bt}$ is a *pseudorandom generator for block size b and space s with parameter δ and seed length ℓ* if, for all branching programs B that read b bits at each step, take t steps, and have width at most 2^s,

$$\left| \Pr_{\gamma \in \{0,1\}^\ell} \big[B(f(\gamma)) \text{ accepts}\big] - \Pr_{\alpha \in \{0,1\}^{bt}} \big[B(\alpha) \text{ accepts}\big] \right| \leq \delta\,.$$

Then Lemma 3 of [6] states the following, in slightly different notation:

Lemma 1. Let $t \leq 2^{n/20}$. Then there is an explicit pseudorandom generator f for block size b and space $b/20$ with parameter $2^{-b/20}$ and seed length $O(b \log t)$.

In our case, to match the union bound over all 2^n possible x we want the parameter δ to be, say, 2^{-2n}. To this end, we modify B_x so that it reads $b = 40n$ bits at each step, ignoring all but n of them. Then [6] gives a pseudorandom generator with seed length $O(n \log t) = O(n \log(1/\varepsilon))$.

Indeed, the space our branching program needs is just $\log(n + m + 1) = O(\log(n/\varepsilon))$, far smaller than the $b/20 = \Theta(n)$ allowed by the lemma. Moreover, if we think of B_x as computing $|w_x| \bmod (n + m + 1)$ (note that $|w_x|$ will never actually wrap around) it becomes a permutation branching program. Furthermore, for uniform inputs we know the probability distribution on the program's states exactly, namely the binomial distribution.

It is tempting to think that these facts allow us to reduce the randomness still further, say to $O(n+\log(1/\varepsilon))$. However, to our knowledge even the best known derandomization results on branching programs under various assumptions [7, 8, 9, 10] require $\Omega((\log 1/\delta)(\log t))$ random bits, even for constant width. Since $\delta = 2^{-\Omega(n)}$ and $t = \Omega(1/\varepsilon^2)$, this again gives $\Omega(n \log(1/\varepsilon))$ random bits.

3 A construction using the Legendre symbol and Weil sums

Here we present another construction, which uses $O(n \log(n/\varepsilon))$ random bits. If ε is fairly large, say $\varepsilon = 1/n^{o(1)}$, then this uses $O(\log n)$ more randomness than the previous construction. However, it is elementary and extremely explicit, and lets us invoke some pretty algebra.

First we recall the definition of the Legendre symbol. Given a prime q, let g be a primitive root, i.e., a multiplicative generator of $\mathbb{F}_q^\times$. Then let $\chi : \mathbb{F}_q \to \mathbb{R}$ be defined as follows:

$$\chi(x) = \begin{cases} +1 & \text{if } x = z^2 \text{ for some } z \neq 0 \\ -1 & \text{if } x = gz^2 \text{ for some } z \neq 0 \\ 0 & \text{if } x = 0\,, \end{cases}$$

This is the quadratic multiplicative character of $\mathbb{F}_q^\times$, extended to $\mathbb{F}_q$ by setting $\chi(0) = 0$. Thus $\chi(xy) = \chi(x)\chi(y)$ for all $x, y \in \mathbb{F}_q$.

Alon, Goldreich, Håstad, and Peralta [2] used the Legendre symbol to construct an ε-biased set as follows. For each $x \in \mathbb{F}_q$, consider the sequence

$$w(x) = (\chi(x+1), \chi(x+2), \ldots, \chi(x+n))\ .$$

Mapping $\{\pm 1\}$ to $\{0, 1\}$ gives an element $\mathbb{F}_2^n$; if $x + i = 0$, we define $w(x)_i = 1$. Their set is then

$$S = \{w(x) \mid x \in \mathbb{F}_q\}\,.$$

Except for a small error due to the rare case where $x + i = 0$, the bias of S with respect to $T \subseteq [n]$ is then

$$b_T = \mathop{\mathbb{E}}_{x \in \mathbb{F}_q} \phi_T(w(x)) = \mathop{\mathbb{E}}_{x \in \mathbb{F}_q} \prod_{i \in T} w(x)_i = \mathop{\mathbb{E}}_{x \in \mathbb{F}_q} \prod_{i \in T} \chi(x+i) = \mathop{\mathbb{E}}_{x \in \mathbb{F}_q} \chi\left(\prod_{i \in T} (x+i)\right).$$

If we write

$$p(x) = \prod_{i \in T} (x + i)\,,$$

then $p(x)$ is a polynomial of degree $|T| \leq n$. In that case, the bias is a Weil sum, which we can bound using the following classic theorem:

Theorem 1 (Weil)**.** Let $p(x) \in \mathbb{F}_q[x]$ be a non-square polynomial of degree d. Then

$$\left| \mathop{\mathbb{E}}_{x \in \mathbb{F}_q} \chi(p(x)) \right| \leq \frac{d-1}{\sqrt{q}}\,.$$

Since $d \leq n$, the bias is bounded by $|b_T| \leq n/\sqrt{q}$. This gives an ε-biased set S of size $q = n^2/\varepsilon^2$.

Our approach is to "re-randomize" this construction. Rather than taking n consecutive Legendre symbols, we let $\Sigma = (s_1, \ldots, s_n) \in F_q^n$ be a collection of n "shifts." For each $x \in \mathbb{F}_q$, these shifts let us extract n bits from the Legendre symbol sequence, giving a string

$$w(x) = (\chi(x + s_1), \chi(x + s_2), \ldots, \chi(x + s_n))\,.$$

In return for choosing these shifts randomly, we get to use a field $\mathbb{F}_q$ considerably larger than the set itself. We then show that S is ε-biased with high probability in Σ by using Theorem 1 to control high moments of the bias.

Let $X \subseteq \mathbb{F}_q$ be an arbitrary set of size ℓ, such as $\{1, \ldots, \ell\}$. Letting x range over X yields a set

$$S = \{w(x) \mid x \in X\} \subseteq \mathbb{F}_2^n\,,$$

with $|S| = \ell$. Assume for now that $x + s_j \neq 0$ for all $x \in X$ and all $j \in [n]$. Then the bias S with respect to $T \subseteq [n]$ is

$$b_T = \left| \mathop{\mathbb{E}}_{x \in X} \phi_T(w(x)) \right| = \left| \mathop{\mathbb{E}}_{x \in X} \prod_{j \in T} \chi(x - s_j) \right|.$$

We will show that, with high probability in Σ, this bias is small for all $T \neq \varnothing$. To this end, we bound its $2k$th moment for some k to be determined below. Expanding its $2k$th power gives products of the form

$$\prod_{t=1}^{2k} \prod_{j \in T} \chi(x_t - s_j)\,, \tag{2}$$

averaged over all tuples $\{x_1, \ldots, x_{2k}\} \in X^{2k}$. For each $x \in X$, let $N(x)$ be the number of times that x appears in this tuple. If $N(x)$ is even for all x, then this product is a square, and is 1 regardless of the s_j. Taking the union bound over all $(2k-1)!! = (2k-1)(2k-3)\cdots 3 \cdot 1$ perfect matchings of $2k$ objects, the probability that this occurs—given that all ℓ^{2k} tuples $\{x_1, \ldots, x_{2k}\}$ are equally likely—is at most

$$\frac{(2k-1)!!}{\ell^k} = \frac{(2k)!}{2^k k!\, \ell^k} \leq \sqrt{2} \left(\frac{2k}{\mathrm{e}\ell}\right)^k ,$$

where we used a form of Stirling's inequality.

On the other hand, if $N(x)$ is odd for some $x \in X$, the product (2) can be written

$$\prod_{j \in T} \chi(p_{x_1,\ldots,x_{2k}}(s_j)) ,$$

where

$$p_{x_1,\ldots,x_{2k}}(s) = \prod_{x : N(x) \text{ odd}} (x - s)$$

is a polynomial of degree at most $2k$. In that case, since the s_j are independent and uniform in $\mathbb{F}_q$, Theorem 1 gives

$$\left| \mathop{\mathbb{E}}_{\Sigma} \prod_{j \in T} \chi(p_{x_1,\ldots,x_{2k}}(s_j)) \right| = \left| \mathop{\mathbb{E}}_{s \in \mathbb{F}_q} \chi(p_{x_1,\ldots,x_{2k}}(s)) \right|^{|T|} \leq \left(\frac{2k-1}{\sqrt{q}}\right)^{|T|} .$$

Putting this all together, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\Sigma} b_T^{2k} &= \mathop{\mathbb{E}}_{\Sigma} \left(\mathop{\mathbb{E}}_{x \in X} \prod_{j \in T} \chi(x_i - s_j) \right)^{2k} \\
&= \mathop{\mathbb{E}}_{x_1,\ldots,x_{2k}} \mathop{\mathbb{E}}_{\Sigma} \prod_{t=1}^{2k} \prod_{j \in T} \chi(x_t - s_j) \\
&\leq \Pr_{\{x_1,\ldots,x_{2k}\}} [N(x) \text{ even for all } x] \\
&\quad + \mathop{\mathbb{E}}_{x_1,\ldots,x_{2k}} \left[\mathop{\mathbb{E}}_{\Sigma} \prod_{t=1}^{2k} \prod_{j \in T} \chi(x_t - s_j) \;\middle|\; N(x) \text{ odd for some } x \right] \\
&\leq \sqrt{2} \left(\frac{2k}{\mathrm{e}\ell}\right)^k + \left(\frac{2k}{\sqrt{q}}\right)^{|T|} . \qquad (3)
\end{aligned}$$

Markov's inequality gives

$$\Pr[|b_T| > \varepsilon] = \Pr[b_T^{2k} > \varepsilon^{2k}] \leq \frac{\mathbb{E}_\Sigma\, b_T^{2k}}{\varepsilon^{2k}} \qquad (4)$$

We now set $q = 4(\mathrm{e}\ell)^2$, making the field quadratically larger than $|S| = \ell$. We also set $k = |T|$, using the $2k$th moment to control parities of weight k. Then combining (3) and (4) gives, for any $|T| \geq 1$,

$$\Pr[|b_T| > \varepsilon] \leq 2\left(\frac{2|T|}{\mathrm{e}\ell\varepsilon^2}\right)^{|T|}.$$

Taking a union bound over all $T \neq \varnothing$ and using $\binom{n}{|T|} \leq (\mathrm{e}n/|T|)^{|T|}$, the probability that any nontrivial parity has bias greater than ε is at most

$$\sum_{T\neq\varnothing} \Pr[|b_T| > \varepsilon] \leq 2\sum_{|T|=1}^{n} \binom{n}{|T|}\left(\frac{2|T|}{\mathrm{e}\ell\varepsilon^2}\right)^{|T|} \leq 2\sum_{|T|=1}^{n}\left(\frac{2n}{\ell\varepsilon^2}\right)^{|T|}. \qquad (5)$$

If we set

$$\ell = \frac{6n}{\delta\varepsilon^2},$$

where $\delta \leq 1$, then bounding (5) with a geometric series gives

$$\sum_{T\neq\varnothing} \Pr[|b_T| > \varepsilon] \leq \frac{2\delta/3}{1-\delta/3} \leq \delta,$$

so the set S is ε-biased with probability $1-\delta$. Finally, our assumption that $x + s_j \neq 0$ for all $x \in X$ and all $j \in [n]$ holds with probability $1 - n\ell/q = 1 - O(\delta\varepsilon^2)$.

How much randomness do we need for this construction? We have to select the shifts $s_1, \ldots, s_n$ independently and uniformly from $\mathbb{F}_q$, and

$$q = 4(\mathrm{e}\ell)^2 = O\left(\frac{n^2}{\varepsilon^4}\right).$$

Thus the number of random bits we need is

$$n\log q = O\big(n\log(n/\varepsilon)\big).$$

4 Further derandomization?

Can we do better? Our approach has a natural barrier at n random bits; since we take a union bound over all 2^n index sets T, we need a probability space of size at least 2^n. Thus any further derandomization, say to $o(n)$ random bits, would have to bound the bias for many parities simultaneously.

The situation is similar for constructing optimal Ramsey graphs, i.e., edge-colored complete graphs on n vertices such that the largest monochromatic clique has size less than $k = 2\log n$. As pointed out in [11], we can do this by choosing the coloring from a $\binom{k}{2}$-wise ε-biased distribution, i.e., a family of functions from the set of edges to $\{0,1\}$ such that the parity of any set of $\binom{k}{2}$ or

fewer edges is odd with probability ε-close to $1/2$. If $\varepsilon = 2^{-k^2}$, the probability that a given clique of size k is monochromatic is $o\big(1/\binom{n}{k}\big)$. So, by the union bound, with high probability there are no monochromatic cliques of size k.

We can generate such families [2] with $O\big(\log\log n + \binom{k}{2} + \log\varepsilon^{-1}\big) = O(\log^2 n)$ random bits. Since we need a probability space of size $\binom{n}{k} = 2^{\Omega(\log^2 n)}$ for the union bound over all $\binom{n}{k}$ cliques to work, this is tight—unless we can do better than the union bound, ensuring simultaneously that many cliques are bichromatic. It is an interesting open question whether this can be reduced to, say, $O(\log n)$ random bits, in which case there are explicit graph families of polynomial size guaranteed to consist largely of optimal Ramsey graphs.

Acknowledgments. We thank Avi Wigderson and Mark Braverman for helpful conversations, and Michał Kotowski for catching several typos. This work was supported by NSF grant CCF-1117426 and ARO contract W911NF-04-R-0009. We are also grateful to the Scholium Project, and particularly the Androkteinos, for inspiration.

References

[1] J. Naor and M. Naor, "Small-bias probability spaces: Efficient constructions and applications." *SIAM Journal on Computing* 22(4):838–856, 1993.

[2] N. Alon, O. Goldreich, J. Håstad, and R. Peralta, "Simple constructions of almost k-wise independent random variables." *Random Structures and Algorithms* 3(3):289–303, 1992.

[3] Yossi Azar, Rajeev Motwani, and Joseph Naor, "Approximating Probability Distributions Using Small Sample Spaces." *Combinatorica* 18(2): 151-171 (1998).

[4] Mahdi Cheraghchi, Amin Shokrollahi, and Avi Wigderson, "Computational Hardness and Explicit Constructions of Error Correcting Codes." Proc. Allerton Conference, 2006, `http://infoscience.epfl.ch/record/101078/files/final.pdf`.

[5] Avraham Ben-Aroya and Amnon Ta-Shma, "Constructing Small-Bias Sets from Algebraic-Geometric Codes." Proc. FOCS, 191–197, 2009.

[6] Noam Nisan, "Pseudorandom Generators for Space-Bounded Computation." *Combinatorica* 12(4):449–461, 1992.

[7] Ran Raz and Omer Reingold, "On Recycling the Randomness of States in Space Bounded Computation." *Proc. STOC* 159–168 (1999).

[8] Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff, "Pseudorandom Generators for Regular Branching Programs." *P*roc. FOCS 40–47 (2010).

[9] Anindya De, "Pseudorandomness for permutation and regular branching programs." *Proc. IEEE Conference on Computational Complexity* 221–231, 2011.

[10] Thomas Steinke, "Pseudorandomness for Permutation Branching Programs Without the Group Theory." *Electronic Colloquium on Computational Complexity* 19: 83 (2012).

[11] M. Naor, "Constructing Ramsey graphs from small probability spaces." IBM Research Report, 1992, `http://www.wisdom.weizmann.ac.il/~naor/PAPERS/ramsey.ps`